DOCUMENT RESUME

ED 320 966 TM 015 292

AUTHOR Kubiak, Anna T.; Cowell, William R.
TITLE Using Multiple DIF Statistics with the Same Items Appearing in Different Test Forms.
PUB DATE Apr 90
NOTE 20p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 17-19, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Difficulty Level; Elementary Secondary Education; *Equated Scores; *Item Bias; Microcomputers; National Programs; Racial Differences; Sex Differences; *Test Format; Testing Programs; Test Items
IDENTIFIERS Educational Testing Service; *Mantel Haenszel Procedure

ABSTRACT

A procedure used to average several Mantel-Haenszel delta difference values for an item is described and evaluated. The differential item functioning (DIF) procedure used by the Educational Testing Service (ETS) is based on the Mantel-Haenszel statistical technique for studying matched groups. It is standard procedure at ETS to analyze test items for their differential difficulty with regard to gender and race/ethnicity of examinees. Test items used to perform the averaging procedure for this study were taken from the item bank of a large national testing program. Each form used in this study consisted of 80 items. Forms selected for the study had a large number of items in common. The delta difference average is the mean of single delta difference statistics across forms weighted inversely to their respective error variances. Overall, the results of differential item functioning analyses are similar to the weighted average procedure. Even if test forms were constructed to have items in common, as is the case in common item equating, the cost of test development and subsequent statistical analyses would be considerable. The weighted average procedure, on the other hand, is quite easy to perform, particularly with the use of a personal computer, and it does not require reformatting of the original data. Seven data tables are included. (TJH)

Using Multiple DIF Statistics With the Same Items Appearing in Different Test Forms

Anna T. Kubiak & William R. Cowell

Educational Testing Service

National Council on Measurement in Education

Boston, 1990

The authors thank Paul Holland for his helpful comments, Lois Worthington and Louann Benton for performing DIF analyses, and Bob Smith and David Wright for reviewing the paper.

Objectives

It is a standard procedure at ETS to analyze test items for their differential difficulty with regard to gender and race/ethnicity of test takers. The differential item functioning (DIF) procedure (Holland & Thayer, 1986) used at ETS is based on the Mantel-Haenszel statistical technique (Mantel & Haenszel, 1959) for studying matched groups. The Mantel-Haenszel method compares the odds of two gender or racial/ethnic groups answering a test item correctly when the members have been matched on the basis of test scores. The DIF indices aid in identifying differences in difficulty caused by characteristics of the test item itself, after real differences in pertinent knowledge and skills have been taken into account by the process of matching.

Groups which are identified for the purpose of DIF analysis are referred to as reference (generally males and Whites) and focal (generally females, Blacks, Asians, Hispanics, etc.). Members of the reference and focal groups are divided into subgroups based on the total test score. Then, the odds that a reference group member answers the item correctly are calculated for every matched subgroup. This calculation is repeated for members of every matched focal group. The next step is to compute the ratio of the odds for the focal group to the odds for the reference group for each matched subgroup. Then, these odds ratios are averaged across the entire score scale, weighted according to the number of individuals at each score value. The estimate resulting from this

procedure is the average factor by which the odds that a reference group member answers the item correctly exceeds the corresponding odds for comparable focal group members. For example, if the odds-ratio value is equal to 1, then both subgroups are equally likely to answer the item correctly or, to state it differently, the item is equally difficult for both subgroups. The MH D-DIF (Mantel-Haenszel delta difference) index, commonly used at ETS, is a scaled odds-ratio. The scale is the ETS delta scale used in test construction and analysis. Delta is a transformation of percent correct to a scale for item difficulty ranging to about 6 for very easy items to about 20 for very difficult items. A delta value of 13 corresponds to 50 percent correct. Consequently, if MH D-DIF is equal to -1.00, it means that the item is one delta unit more difficult for the focal group than it is for comparable members of the reference group. Near the middle of the difficulty scale, a difference of 1 delta point is equivalent to about 10 percent difference in percent correct. The interpretation would be reversed for positive values of MH D-DIF.

Based on the absolute value of MH D-DIF statistics and their level of significance (α=.05), test items are classified into three categories. The three categories are labeled A, B, and C. Category A contains items with MH D-DIF between -1 and +1 or MH D-DIF not significantly different from 0. Category B is assigned to items with MH D-DIF significantly different from 0, and an absolute value of at least 1, but not significantly greater than 1. Category C characterizes items

with absolute value of MH D-DIF at least equal to 1.5, and significantly greater than 1.

Each ETS testing program that meets minimum sample size requirements performs DIF analyses. This paper refers to DIF analyses completed ordinarily for test assembly purposes, with sample size requirements for the focal and reference groups combined of 500 or greater, and for the focal group of 100 or greater. Some testing programs administer new forms more than once during the same 12-month testing cycle. As a result, items that appear in multiple forms are analyzed several times and have multiple MH D-DIF statistics and classifications. Since each MH D-DIF is influenced by the difficulty of other items included in any particular form, the magnitude, significance, and classification of each MH D-DIF differs occasionally. Consequently, test developers have to decide which MH D-DIF value to use during the test assembly process. Currently, the choices involve using the first value, the most recent value, the largest (worst case) value, or some average value (such as the median).

The purpose of this paper is to describe a procedure suggested by Paul Holland (personal communication, May 16, 1989) that can be used to average several MH D-DIF values for an item, to test the significance of the average, and to obtain a new classification. This procedure is not exact since each MH D-DIF value for the same item was calculated with different sets of items comprising each test form and, consequently, the criterion score. Therefore, this

averaging is compared with the exact technique that involves combining samples to whom the common items were administered and recalculating MH D-DIF statistics from item response data.

Instruments

Test items used to perform the averaging procedure were taken from the item bank of a large national testing program. Each form used in this study consisted of 80 items. Three pairs of forms that contained overlapping items were selected. Form 7 and Form 11 had 30 common items, Form 4 and Form 6 had 48 items in common, and Form 3 and Form 5 had 52 items in common. These forms were selected for this study because they have a large number of items in common. Ordinarily, a smaller number of items overlaps between forms. Forms with large number of common items were selected for this study in order to provide a meaningful matching criterion score, i.e. in addition to the total test score criterion, the common items score criterion was also used in DIF analyses.

Method

The MH D-DIF average is the mean of single MH D-DIF statistics across forms weighted inversly to their respective error variances. The first step therefore is to calculate weights proportional to the inverse of error variances with the constraint that the sum of the weihgts must be equal to 1, $w_i = k\Sigma\tilde{e}_i^2$ and $\Sigma w_i = 1$. The second step is to compute the MH D-DIF weighted mean, $MHAVG = \Sigma MH\ D\text{-}DIF_i w_i$. In the third step, the standard error of the weighted average is calculated by

taking a square root of the sum of the squared weights (from step 1) multiplied by the original error variances, $SEW=\sqrt{\Sigma w_i^2 SE_i^2}$. The last step is to perform a significance test for the average MH D-DIF using the new standard error and assign a category (A, B, or C) depending on the outcome. The statistical significance was tested using z-test and $\alpha=.05$. All computations were performed with the use of SPSS-X/PC+.

For the second part of this study, standard DIF analyses were performed. Each test form was analyzed twice, first with the total test criterion score, and next with the common items criterion score. These analyses were performed using single samples of examinees who took either of the paired forms. In the next step, DIF analyses were executed using combined samples of candidates to whom each pair of test forms was administered, i.e. examinees who took Form 7 and those who took Form 11 were grouped together for each analysis. Two DIF analyses were run for each pair of forms, one with the total test score as the criterion and second with the common items score as the criterion. Altogether, six analyses were performed for each pair of forms. The following example lists analyses for Forms 7 and 11:

- Form 7, single group, total test score criterion
- Form 7, single group, common items score criterion
- Form 11, single group, total test score criterion
- Form 11, single group, common items score criterion
- Form 7 & 11, combined groups, total test score criterion
- Form 7 & 11, combined groups, common items score criterion.

Results

Table 1 presents sample sizes for male/female comparisons for each test form used in this study. Table 2 and Table 3 present results for Forms 7 and 11, Tables 4 and Table 5 present information for Forms 4 and 6, and Table 6 and Table 7 contain results of analyses for Forms 3 and 5. Each table lists the item number in each form, MH D-DIF values, standard errors, and classifications obtained from separate analyses with single sample sizes. In addition, each table contains MH D-DIF statistics, standard errors, and classifications obtained from analyses performed with combined samples. Finally, each table contains results of the weighted average procedure. The mean values of MH D-DIF statistics and standard errors are listed at the bottom of each table.

Comparisons between the exact procedure which involves combining samples and recalculating DIF statistics, and the weighted average procedure, should be the most straightforward when the common item criterion score is used. In this case, the influence or confounding caused by the "uncommon" items, that is items different in each form which also contributed to the total test score criterion, is eliminated. The following description is focused on the results of analyses with combined samples and on the results of the averaging. The results of single form analyses are listed for reference purposes.

When the criterion score of 30 common items was used to analyze Forms 7 and 11 using combined samples, DIF analyses resulted in two C classifications, for items 3 and 95, and five B classifications,

for items 33, 38, 42, 68, 85. The remaining items were classified as A's. The weighted average procedure that was performed on MH D-DIF values listed in Table 2, resulted in identical classifications for the same items. Table 3 presents the outcomes of analyses with the total test score as the criterion for Forms 7 and 11. Although the comparisons are not as straightforward, this situation is more like that in which the procedure would be applied in a typical testing situation. The combined samples analysis resulted in one C classification, for item 95, and seven B classifications, for items 3, 16, 38, 42, 68, 75, 85. The weighted average procedure resulted in one C classification, for item 95, and five B classifications, for items 16, 38, 42, 75, 85. The two procedures, combined samples and weighted average with total score as the criterion, differed by one classifications for item 68. This item was classified as B with combined samples, but when the averaging procedure was used, it was classified as A. Item 68 obtained MH D-DIF of -1.07 with the combined samples approach, and -.97 with the weighted average method. The two values differed only by .10, but since one of the conditions used for classifying items as B is that the absolute value of their MH D-DIF must be at least 1, MH D-DIF equal to -1.07 was classified as B, whereas MH D-DIF equal to -.97 obtained A.

Tables 4 and 5 contain information about Form 4 and Form 6. These two forms had 48 items in common. Table 4 presents results of the analyses with the 48 items score as the criterion. The combined samples and the weighted average procedures each resulted in ten

B classifications, for items 32, 38, 44, 45, 46, 70, 71, 72, 74, 76. Table 5 contains results of DIF analyses with the total test score as the criterion. The combined samples procedure and the averaging procedure each resulted in eight B classifications, for items 32, 38, 44, 45, 46, 71, 74, 76. Items 70 and 72 were classified as A with the total test score criterion, but as B with the commom items criterion. This difference in classification could be explained by the influence of the unique items that were also included in the total test score criterion.

Tables 6 and 7 present results of DIF analyses for Form 3 and Form 5 with 52 common items. When the common items score criterion was used (Table 6), items 45, 52, 73, 75, 93 were classified as B by both, the combined samples method and the weighted average method. The difference in classification, for items 35 and 64, was due to the convention used for assigning the B category. The value of MH D-DIF must not only be significantly different from 0 but its absolute value must be at least equal to 1. Item 80 obtained MH D-DIF of -1.24 with combined samples which was equal to the critical value at $\alpha = .05$. Therefore, the result of the significance test yielded an A, since the tested value must exceed the critical value for the H_o to be rejected. When the total test score was used as the criterion (Table 7), items 45, 52, 60, 73, 75, 93 were categorized as B by both, the combined samples method and weighted average method. Item 64 obtained borderline MH D-DIF values of -1.11 with combined samples and -.99 when the weighted average was used.

Educational Importance of Study

The technique of averaging MH D-DIF statistics presented in this paper makes use of multiple outcomes of DIF analyses performed on the same items. When multiple administrations of the same items in different test forms occur, the best procedure is to combine all samples to which these items were administered and perform new DIF analysis. This procedure, however, requires additional reformatting of the data files and, therefore, is inefficient and costly. Also, there are seldom enough common items to provide a meaningful matching criterion. If these "combined" results are not available, test developers must choose among multiple MH D-DIF statistics, where the most recent MH D-DIF may be selected as opposed to the "older" ones, or the largest or median value may be used to judge differential performance of items. If the largest or "worst-case" value is systematically selected, number of B and C items will increase as the number of analyses increases. Although the averaging procedure makes use of all MH D-DIF statistics and their standard errors, it is not an exact procedure because the matching criterion, that of total test score, is different each time MH D-DIF is computed (not all items are identical between forms).

Overall, the results of DIF analyses, with combined samples and the common items score as the criterion and the total test score as the criterion, are similar to the weighted average procedure when each of the above criteria was used. However, no statistical test was used to decide whether the difference between each pair of absolute MH D-DIF values was significant. From the practical significance

standpoint, classifications resulting from the averaging procedure were similar enough to classifications obtained from combining samples and recalculating MH D-DIF statistics. In one case (Forms 4 and 6), both procedures resulted in identical classifications for the same items. DIF analysis with the total test criterion score resulted in one different classification for Forms 7 and 11. When Forms 3 and 5 were analyzed with the common items criterion, three items were classified differently by the combined samples method and the weighted average approach. With the total test score criterion, Forms 3 and 5 differed by one classification when these two methods were used. The magnitude of differences between the absolute values of MH D-DIF and their standard errors for the discrepant cases was very small.

Even if test forms were constructed to have items in common, as is the case in common item equating, the cost of test development and subsequent statistical analyses would be considerable. Also, when the total test score is used as the criterion with combined samples, it is not possible, at the present time, to refine the criterion by deleting C items. This problem involves using the total test score for each examinee, as opposed to each person's single responses to each test item. The weighted average procedure, on the other hand, is quite easy to perform especially with the use of a PC, and it does not require reformatting of the original data. However, if the weighted average technique were to be used operationally in other testing programs or for other focal groups, further analyses should be performed with the other reference and focal groups.

References

Holland, P.W., & Thayer, D.T. (1986, April). Differential item performance and the Mantel-Haenszel procedure. Paper presented at the meeting of the American Educational Research Association, San Francisco.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis data from retrospective studies of disease. Journal of the National Cancer Institute, 22, 719-748.

Patrick, R. (1989). IANA83 User's Guide. DIF and Distractor Analysis. Educational Testing Service, Princeton, NJ.

SPSS/PC+ Base Manual (1988). SPSS Inc., Chicago, IL.

TABLE 1
Sample Sizes for Reference and Focal Groups

	FORM 3	FORM 4	FORM 5	FORM 6	FORM 7	FORM 11
MALE	395	299	346	340	378	277
FEMALE	332	240	274	258	316	259
------	---	---	---	---	---	---
TOTAL	727	539	620	598	694	536

TABLE 2
Single, Combined, and Averaged DIF Statistics
Criterion = Common Items Score

	FORM 7			FORM 11				COMBINED			AVERAGE	
ITEM	CLASS*	MH-DDIF	SE	NUMBER	CLASS	MD-DDIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
3	C	-1.90	0.51	43	B	-1.69	0.60	C	-1.76	0.39	C	-1.80	0.39
4	B	-1.16	0.52	39		-0.66	0.58		-0.87	0.39		-0.92	0.39
5		-0.21	0.43	54		0.81	0.48		0.27	0.32		0.27	0.32
7		-0.36	0.45	64		-0.72	0.54		-0.47	0.35		-0.52	0.35
9		-0.65	0.57	33		-0.90	0.64		-0.84	0.42		-0.77	0.43
10		0.45	0.41	82		0.65	0.47		0.52	0.31		0.54	0.31
16		0.73	0.43	1		1.00	0.51		0.83	0.32		0.85	0.33
18		-0.45	0.99	2		-1.53	0.91		-0.96	0.64		-1.01	0.67
26		-1.72	1.49	72		-1.54	1.46		-1.55	1.02		-1.63	1.04
28		-0.03	0.38	87		0.50	0.43		0.15	0.28		0.22	0.29
33	B	-1.21	0.53	48	B	-1.33	0.60	B	-1.19	0.30	B	-1.27	0.40
37		0.33	0.38	92		0.36	0.44		0.30	0.29		0.34	0.29
38	B	1.59	0.57	16	B	1.70	0.63	B	1.54	0.42	B	1.64	0.42
42	B	1.90	0.78	27		0.82	0.78	B	1.27	0.54	B	1.36	0.55
44		0.41	0.51	44		-0.28	0.61		0.10	0.31		0.10	0.39
53		0.55	0.41	88		-0.36	0.46		0.13	0.30		0.12	0.31
60		-0.63	0.56	28		-1.02	0.67		-0.81	0.43		-0.81	0.43
62		-0.26	0.40	90		-0.59	0.46		-0.45	0.30		-0.41	0.30
67		-0.97	0.58	31		-0.52	0.66		-0.76	0.43		-0.76	0.44
68	B	-1.34	0.41	89	B	-1.11	0.46	B	-1.32	0.30	B	-1.23	0.31
75		0.93	0.41	91		0.88	0.46		0.94	0.30		0.91	0.31
81		-0.63	0.49	38		-0.24	0.60		-0.47	0.38		-0.45	0.38
83		-0.49	0.67	67		-0.18	0.86		-0.28	0.52		-0.35	0.53
85	C	2.47	0.71	29		-0.37	0.67	B	1.07	0.47	B	1.01	0.49
86		-0.55	0.38	94		-0.83	0.44		-0.71	0.29		-0.68	0.29
88		0.22	0.67	22		-0.28	0.80		-0.14	0.51		-0.01	0.52
89		-0.06	0.44	63		0.77	0.50		0.26	0.33		0.33	0.33
91	B	1.12	0.47	61		0.64	0.48		0.76	0.34		0.88	0.34
92		0.74	0.44	73		0.72	0.48		0.64	0.32		0.73	0.32
95	C	2.10	0.56	41	B	1.66	0.61	C	1.79	0.41	C	1.89	0.41
		-----	-----			-----	-----		-----	-----		-----	-----
MEAN		-0.42	0.55			-0.47	0.61		-0.42	0.40		-0.42	0.41

*Blank spaces indicate A items

TABLE 3
Single, Combined, and Averaged DIF Statistics
Criterion = Total Test Score

FORM 7				FORM 11				COMBINED			AVERAGE		
ITEM	CLASS	MH-DDIF	SE	ITEM	CLASS	MH-DDIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
3	B	-1.48	0.52	43		1.03	0.64	B	-1.22	0.40	B	-1.28	0.41
4		-0.67	0.53	39		-0.04	0.60		-0.54	0.39		-0.37	0.40
5		0.08	0.44	54		0.66	0.51		0.42	0.33		0.35	0.33
7		0.02	0.46	64		-0.24	0.58		-0.19	0.35		-0.10	0.36
9		-0.35	0.58	33		-0.60	0.68		-0.53	0.43		-0.47	0.44
10		0.71	0.42	82		0.87	0.49		0.70	0.31		0.78	0.32
16		0.94	0.41	1	B	1.58	0.53	B	1.27	0.32	B	1.22	0.33
18		0.52	1.03	2		-1.52	0.94		-0.56	0.67		-0.55	0.70
26		-1.25	1.50	72		-0.84	1.42		-1.08	1.04		-1.04	1.03
28		0.14	0.39	87		0.61	0.44		0.29	0.29		0.36	0.29
33		-0.57	0.53	48		-.04	0.62		-0.69	0.40		-0.79	0.40
37		0.27	0.39	92		0.38	0.45		0.35	0.29		0.32	0.30
38	B	1.65	0.58	16	B	1.80	0.67	B	1.64	0.43	B	1.72	0.44
42	B	1.76	0.80	27		1.16	0.83	B	1.29	0.55	B	1.47	0.58
44		0.81	0.52	44		0.02	0.65		0.48	0.40		0.46	0.41
53		0.83	0.40	88		-0.07	0.47		0.47	0.30		0.42	0.31
60		-0.40	0.57	28		-0.54	0.68		-0.56	0.44		-0.46	0.44
62		-0.02	0.41	90		-0.15	0.48		-0.16	0.31		-0.08	0.31
67		-0.74	0.60	31		-0.22	0.70		-0.64	0.45		-0.50	0.46
68	B	-1.13	0.40	89		-.79	0.46	B	-1.07	0.30		-0.97	0.30
75	B	1.21	0.41	91	B	1.41	0.48	B	1.26	0.31	B	1.30	0.31
81		-0.19	0.50	38		0.17	0.63		-0.15	0.39		-0.03	0.39
83		-0.09	0.66	67		0.14	0.97		0.07	0.54		0.00	0.56
85	C	2.76	0.72	29		0.30	0.74	B	1.74	0.49	B	1.55	0.52
86		-0.61	0.38	94		-0.88	0.46		-0.65	0.29		-0.73	0.[illegible]
88		0.61	0.70	22		0.84	0.81		0.03	0.52		0.72	0.53
89		-0.22	0.44	63		0.44	0.52		0.24	0.33		0.08	0.34
91		0.93	0.48	61		0.24	0.50		0.69	0.34		0.59	0.35
92		0.80	0.43	73		0.86	0.50		0.85	0.32		0.83	0.33
95	C	2.12	0.56	41	B	1.83	0.62	C	1.94	0.41	C	1.98	0.42
		.....	.....			.....	.....		.....	.....		.....	.....
MEAN		-0.26	0.56			-0.[illegible]3	0.64		-0.27	0.41		-0.25	0.42

TABLE 4
Single, Combined, and Averaged DIF Statistics
Criterion = Common Items Score

FORM 4				FORM 6				COMBINED			AVERAGE		
ITEM	CLASS	MH-DDIF	SE	ITEM	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
2		-0.01	0.52	63		-0.61	0.47		-0.37	0.34		-0.33	0.35
4		-0.60	0.47	91		-0.80	0.46		-0.81	0.33		-0.70	0.33
7		0.00	0.55	14	B	1.93	0.63		0.89	0.40		0.90	0.42
8		0.50	0.44	48		0.16	0.42		0.31	0.30		0.33	0.30
11		0.19	0.49	7		0.21	0.48		0.22	0.34		0.20	0.34
15		0.85	0.61	80		0.77	0.58		0.77	0.41		0.81	0.42
16		-0.52	0.51	71		-0.59	0.48		-0.56	0.34		-0.56	0.35
17		0.32	0.51	5		0.44	0.46		0.29	0.33		0.38	0.34
18		-0.28	0.58	4		0.24	0.57		-0.06	0.40		-0.02	0.41
24		-0.48	0.46	66		-0.36	0.43		-0.45	0.31		-0.42	0.31
27		-0.32	0.50	6		0.11	0.44		-0.07	0.33		-0.09	0.33
30		-0.61	0.53	12		0.82	0.53		0.13	0.37		0.10	0.37
32		1.32	0.70	9		1.09	0.60	B	1.10	0.45	B	1.20	0.46
38	B	-1.09	0.49	2	B	-1.17	0.44	B	-1.11	0.32	B	-1.13	0.33
40		-0.19	0.61	26		1.33	0.70		0.43	0.45		0.52	0.46
44		-1.40	0.69	78		-0.85	0.60	B	-1.08	0.44	B	-1.11	0.45
45	B	1.60	0.67	25	B	1.46	0.60	B	1.61	0.44	B	1.53	0.45
/		1.19	1.11	87		1.66	0.89	B	1.46	0.68	B	1.45	0.70
49		0.34	0.48	69		0.04	0.44		0.12	0.32		0.18	0.32
50	B	1.40	0.55	19		0.16	0.50		0.60	0.36		0.75	0.37
51		-0.43	0.82	29		-0.55	0.75		-0.56	0.55		-0.49	0.55
52	B	-1.20	0.46	30		-0.49	0.43		-0.80	0.30		-0.83	0.31
53		-0.57	0.59	84		-0.20	0.46		-0.33	0.36		-0.36	0.37
54		0.74	0.45	85		0.15	0.42		0.44	0.30		0.43	0.31
55		0.21	0.48	81		-0.27	0.45		0.00	0.32		-0.04	0.33
58		0.34	0.49	33	B	1.09	0.49		0.74	0.35		0.72	0.35
59		-0.39	0.60	34		-0.29	0.58		-0.36	0.42		-0.34	0.42
61		-0.51	0.90	20		0.27	0.71		0.12	0.55		-0.07	0.56
64		0.38	0.44	36		-0.19	0.43		-0.01	0.30		0.09	0.31
66		-0.21	0.52	86		-0.25	0.44		-0.11	0.33		-0.23	0.34
68		0.20	0.51	90		-0.29	0.50		-0.05	0.35		-0.05	0.36
70	B	1.97	0.85	92		0.63	0.65	B	1.06	0.50	B	1.21	0.52
71	B	1.57	0.66	54		1.10	0.56	B	1.38	0.42	B	1.32	0.43
72		-0.89	0.60	89		-1.39	0.62	B	-1.08	0.42	B	-1.14	0.43
74	B	-1.44	0.51	70		-0.67	0.49	B	-1.20	0.35	B	-1.05	0.35
75		0.01	0.57	57		0.13	0.52		0.07	0.38		0.07	0.38
76	C	3.85	1.18	47		0.05	0.77	B	1.28	0.61	B	1.55	0.64
78		-0.51	0.54	67		-0.21	0.50		-0.31	0.36		-0.35	0.37
79	B	-1.03	0.48	49		-0.10	0.44		-0.58	0.32		-0.54	0.32
80		-0.45	0.46	62		-0.41	0.43		-0.43	0.31		-0.43	0.31
84		0.54	0.57	59		0.00	0.48		0.21	0.36		0.25	0.37
85	B	1.08	0.49	61		0.07	0.49		0.53	0.34		0.51	0.35
86		0.68	0.53	15		-0.38	0.60		0.18	0.39		0.18	0.40
87		-0.37	0.50	53		-0.17	0.45		-0.13	0.33		-0.26	0.33
89		-0.33	0.46	65		0.35	0.43		0.10	0.31		0.02	0.31
92		0.30	0.46	55		-0.51	0.44		-0.14	0.32		-0.11	0.32
93		0.50	0.43	8		0.82	0.43		0.59	0.30		0.66	0.30
94		0.11	0.50	52		0.76	0.46		-0.32	0.33		0.45	0.34
		-----	-----			-----	-----		-----	-----		-----	-----
MEAN		-0.29	0.57			-0.23	0.52		-0.23	0.38		-0.22	0.39

TABLE 5
Single, Combined, and Averaged DIF Statistics
Criterion = Total Test Score

FORM 4				FORM 6				COMBINED			AVERAGE		
ITEM	CLASS	MH-DDIF	SE	ITEM	CLASS	MD-DDIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
2		-0.06	0.52	63		-0.65	0.48		-0.25	0.48		-0.37	0.35
4		-0.66	0.47	91		-0.86	0.47		-0.75	0.33		-0.76	0.33
7		-0.05	0.56	14	B	2.06	0.68		0.79	0.41		0.90	0.43
8		0.36	0.44	48		0.14	0.43		0.23	0.30		0.25	0.31
11		0.19	0.49	7		0.04	0.49		0.23	0.34		0.12	0.35
15		0.87	0.63	80		0.50	0.60		0.62	0.42		0.68	0.43
16		-0.26	0.52	71		-0.53	0.48		-0.42	0.35		-0.40	0.35
17		0.18	0.51	5		0.55	0.47		0.40	0.33		0.37	0.35
18		-0.42	0.58	4		0.23	0.59		-0.01	0.40		-0.10	0.41
24		-0.55	0.46	66		-0.50	0.43		-0.48	0.31		-0.52	0.31
27		-0.17	0.49	6		0.14	0.45		0.01	0.32		-0.01	0.33
30		-0.20	0.54	12		0.71	0.54		0.22	0.37		0.26	0.38
32	B	1.38	0.69	9	B	1.18	0.62	B	1.24	0.45	B	1.27	0.46
38		-0.96	0.48	2	B	-1.20	0.45	B	-1.11	0.32	B	-1.08	0.33
40		-0.39	0.62	26		1.31	0.70		0.35	0.44		0.41	0.46
44		-1.46	0.69	78		-0.60	0.63	B	-1.00	0.45	B	-1.01	0.47
45	B	1.75	0.67	25	B	1.32	0.64	B	1.55	0.44	B	1.53	0.46
46		1.67	1.07	87		1.94	1.06	B	1.83	0.74	B	1.81	0.75
49		0.30	0.48	69		-0.03	0.45		0.18	0.32		0.13	0.33
50	B	1.11	0.53	19		0.04	0.50		0.58	0.36		0.56	0.36
51		-0.05	0.81	29		-0.52	0.75		-0.31	0.54		-0.29	0.55
52	B	-1.00	0.46	30		-0.47	0.43		-0.75	0.31		-0.73	0.31
53		-0.47	0.48	84		-0.20	0.47		-0.27	0.36		-0.33	0.34
54		0.81	0.46	85		0.07	0.43		0.31	0.30		0.43	0.31
55		0.46	0.48	81		-0.22	0.46		0.00	0.32		0.11	0.33
58		-0.26	0.18	33		0.98	0.50		0.62	0.34		0.45	0.19
59		-0.30	0.12	34		-0.18	0.61		-0.30	0.42		-0.28	0.14
61		0.13	0.87	20		0.34	0.74		0.36	0.55		0.24	0.57
64		0.25	0.22	36		-0.19	0.44		-0.06	0.31		0.10	0.21
66		-0.03	0.53	86		-0.24	0.45		-0.10	0.33		-0.14	0.34
68		0.06	0.51	90		-0.50	0.50		-0.14	0.35		-0.22	0.36
70	B	1.46	0.82	92		0.31	0.69		0.92	0.51		0.84	0.53
71	B	1.55	0.65	54	B	1.45	0.61	B	1.38	0.43	B	1.50	0.45
72		-0.74	0.59	89		-1.15	0.62		-0.89	0.42		-0.94	0.43
74	B	-1.74	0.54	70		-0.57	0.51	B	-1.04	0.35	B	-1.14	0.37
75		0.06	0.58	57		0.14	0.54		0.07	0.38		0.10	0.40
76	C	3.00	1.25	47		0.31	0.83	B	1.37	0.63	B	1.56	0.71
78		-0.55	0.77	67		0.00	0.51		-0.35	0.37		-0.22	0.43
79	B	-1.03	0.48	49		-0.27	0.45		-0.59	0.32		-0.64	0.33
80		-0.29	0.45	62		-0.21	0.43		-0.33	0.31		-0.25	0.31
84		0.41	0.57	59		0.06	0.49		0.27	0.36		0.22	0.37
85	B	1.57	0.51	61		0.01	0.49		0.66	0.34		0.77	0.35
86		0.62	0.54	15		-0.58	0.62		0.10	0.40		0.06	0.41
87		0.05	0.50	53		-0.04	0.46		-0.05	0.33		0.00	0.34
89		-0.30	0.46	65		0.51	0.44		0.14	0.31		0.11	0.32
92		0.50	0.48	55		-0.29	0.46		-0.06	0.32		0.10	0.33
93		0.30	0.44	8		0.55	0.43		0.45	0.30		0.43	0.31
94		0.17	0.50	52		-0.86	0.48		-0.28	0.34		-0.36	0.35
		-----	-----			-----	-----		-----	-----		-----	-----
MEAN		-0.24	0.56			-0.23	0.54		-0.20	0.38		-0.20	0.38

TABLE 6

Single, Combined, and Averaged DIF Statistics

Criterion = Common Items Score

	FORM 3			FORM 5				COMBINED			AVERAGE	
ITEM	CLASS	MH-DDIF	SE	ITEM	CLASS	MH-DDIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
3		0.38	0.48	94		0.21	0.53		0.38	0.35		0.30	0.36
7		-0.15	0.57	28		0.07	0.58		0.75	0.28		-0.04	0.41
9		0.31	0.38	49		0.32	0.43		0.37	0.28		0.31	0.29
12		-0.26	0.42	39		0.57	0.44		0.13	0.30		0.15	0.30
13		0.29	0.57	68		-1.04	0.62		-0.26	0.41		-0.35	0.42
14		-0.33	0.39	59	B	-1.40	0.42		-0.86	0.28		-0.85	0.29
21		0.30	0.42	35		0.42	0.45		0.39	0.31		0.36	0.31
25		0.25	0.42	51		-0.14	0.42		0.01	0.29		0.05	0.30
26		-0.25	0.49	56		-0.64	0.50		-0.44	0.35		-0.44	0.33
28		0.38	0.48	31		-0.35	0.50		0.05	0.34		0.02	0.35
29		-0.12	0.54	2		-0.39	0.51		-0.34	0.37		-0.26	0.37
30		-0.41	0.40	18		-0.30	0.43		-0.38	0.29		-0.36	0.29
31		-0.16	0.41	90		0.63	0.43		0.25	0.30		0.23	0.30
32		0.28	0.39	45		-0.12	0.41		0.08	0.28		0.09	0.28
34		0.07	0.45	86		-0.86	0.46		-0.38	0.32		-0.39	0.32
35	B	-1.46	0.59	82		-0.51	0.57	B	-1.01	0.41		-0.98	0.41
37		-0.28	0.39	8		0.16	0.42		-0.13	0.28		-0.07	0.29
39		-0.18	0.41	70		0.32	0.44		0.04	0.30		0.06	0.30
40		0.04	0.43	95		0.22	0.45		0.11	0.31		0.13	0.31
45		0.94	0.53	85	B	1.46	0.56	B	1.18	0.38	B	1.19	0.39
46		-0.15	0.39	61		-0.27	0.43		-0.18	0.28		-0.21	0.29
48		0.90	0.55	73		0.07	0.74		0.50	0.43		0.55	0.45
49		0.40	0.40	69	B	1.02	0.43		0.70	0.29		0.70	0.29
51		-0.22	0.41	67		-0.09	0.45		-0.17	0.30		-0.16	0.30
52	B	-1.01	0.43	23	B	-1.74	0.46	B	-1.31	0.31	B	-1.36	0.31
54		-1.37	0.83	7		0.25	0.73		-0.30	0.53		-0.51	0.55
60		0.22	0.61	58		1.27	0.76		0.68	0.47		0.69	0.48
62		0.12	0.42	14		0.82	0.48		0.47	0.31		0.45	0.32
64	B	-1.22	0.60	25		-0.72	0.57	B	-1.02	0.41		-0.96	0.41
65		0.37	0.48	75		0.30	0.49		0.32	0.34		0.34	0.34
68		0.10	0.43	66		0.07	0.48		0.01	0.32		0.09	0.32
69		0.52	0.47	48		0.80	0.52		0.68	0.35		0.65	0.35
70		-0.62	0.40	20	B	-1.36	0.46		-0.92	0.30		-0.96	0.30
72		0.17	0.38	42		0.47	0.40		0.27	0.27		0.32	0.28
73	B	2.45	1.15	22		1.73	0.98	B	2.10	0.74	B	2.06	0.75
74		0.31	0.41	76	B	1.04	0.44		0.67	0.30		0.66	0.30
75	B	1.15	0.58	79	B	1.89	0.73	B	1.51	0.45	B	1.48	0.46
77		0.66	0.40	37		-0.29	0.45		0.24	0.29		0.21	0.30
78		-0.35	0.37	5		-0.08	0.40		-0.26	0.27		-0.22	0.27
80		-1.14	0.84	83		-1.44	0.96		-1.24	0.63	B	-1.28	0.63
81		1.10	0.82	36		0.65	0.82		1.01	0.58		0.88	0.58
83		-0.29	0.45	3		0.80	0.47		0.24	0.32		0.24	0.33
84		0.39	0.38	30		-0.28	0.42		0.10	0.28		0.07	0.28
85		-0.03	1.36	32		-0.38	1.35		-0.47	0.95		-0.21	0.96
86		0.52	0.62	92		-0.49	0.73		0.02	0.47		0.06	0.47
87		0.76	0.39	27		0.66	0.41		0.75	0.23		0.71	0.28
88		-1.23	0.71	87		-0.27	0.68		-0.81	0.48		-0.74	0.49
89		0.45	0.50	44		0.24	0.58		0.34	0.38		0.35	0.38
91		-0.45	0.45	24		-0.26	0.50		-0.37	0.34		-0.36	0.33
92		-0.26	0.63	1		0.27	0.63		-0.04	0.44		0.01	0.45
93		-0.61	0.67	38	B	-1.44	0.66	B	-1.02	0.46	B	-1.03	0.47
94		-0.93	0.39	84		-0.89	0.42		-0.93	0.28		-0.91	0 29
		-----	-----			-----	-----		-----	-----		-----	-----
MEAN		-0.26	0.52			-0.30	0.55		-0.25	0.37		-0.24	0.38

TABLE 7

Single, Combined, and Averaged DIF Statistics

Criterion = Total Test Score

FORM 3				FORM 5				COMBINED			AVERAGE		
ITEM	CLASS	MH-DDIF	SE	ITEM	CLASS	MD-DDIF	SE	CLASS	MH D-DIF	SE	CLASS	MH D-DIF	SE
3		0.41	0.50	94		0.25	0.54		0.43	0.36		0.33	0.37
7		-0.38	0.57	28		-0.02	0.57		-0.16	0.40		-0.20	0.40
9		0.42	0.39	49		0.48	0.43		0.42	0.29		0.45	0.29
12		-0.04	0.43	39		0.74	0.45		0.25	0.31		0.34	0.31
13		0.12	0.61	68		-0.60	0.64		-0.20	0.42		-0.23	0.44
14		-0.23	0.41	59	B	-1.29	0.43		-0.75	0.29		-0.75	0.29
21		0.44	0.45	35		0.49	0.46		0.50	0.32		0.46	0.32
25		0.05	0.43	51		-0.10	0.42		-0.02	0.30		-0.03	0.30
26		-0.11	0.50	56		-0.32	0.49		-0.21	0.35		-0.22	0.35
28		0.46	0.50	31		-0.25	0.50		0.07	0.34		0.11	0.35
29		-0.68	0.57	2		-0.45	0.52		-0.52	0.38		-0.56	0.38
30		-0.35	0.41	18		-0.13	0.45		-0.26	0.30		-0.25	0.30
31		-0.17	0.42	90		0.47	0.44		0.21	0.30		0.14	0.30
32		0.27	0.40	45		0.07	0.41		0.19	0.28		0.17	0.29
34		-0.21	0.47	86		-0.76	0.46		-0.35	0.32		-0.49	0.33
35	B	-1.29	0.60	82		-0.50	0.59		-0.83	0.41		-0.89	0.42
37		-0.35	0.40	8		0.32	0.42		0.00	0.29		-0.02	0.29
39		-0.09	0.41	70		0.26	0.44		0.05	0.30		0.08	0.30
40		-0.08	0.43	95		0.32	0.46		0.17	0.31		0.11	0.31
45		0.92	0.54	85	B	1.51	0.56	B	1.09	0.38	B	1.21	0.39
46		-0.22	0.40	61		-0.01	0.43		-0.16	0.29		-0.12	0.29
48		0.82	0.58	73		0.28	0.73		0.45	0.44		0.58	0.46
49		0.37	0.41	69	B	1.00	0.44		0.70	0.29		0.67	0.30
51		-0.25	0.41	67		-0.06	0.45		-0.23	0.30		-0.16	0.30
52	B	-1.17	0.44	23	B	-1.69	0.48	B	-1.41	0.32	B	-1.42	0.32
54		-0.79	0.83	7		0.45	0.74		-0.14	0.52		-0.13	0.55
60		0.80	0.64	58		1.45	0.76		0.97	0.47	B	1.10	0.49
62		0.13	0.43	14	B	1.01	0.48		0.49	0.31		0.55	0.32
64		-1.22	0.62	25		-0.77	0.59	B	-1.11	0.42		-0.99	0.43
65		0.21	0.49	75		0.45	0.50		0.27	0.34		0.33	0.35
68		0.22	0.44	66		0.26	0.47		0.14	0.32		0.24	0.32
69		0.30	0.49	48		1.02	0.53		0.61	0.35		0.65	0.36
70		-0.57	0.43	20	B	-1.26	0.46		-0.80	0.31		-0.90	0.31
72		0.17	0.38	42		0.57	0.41		0.36	0.28		0.36	0.28
73		1.84	1.23	22		1.50	1.02	B	1.83	0.77	B	1.65	0.79
74		0.29	0.42	76	B	1.23	0.44		0.77	0.30		0.75	0.30
75		1.17	0.62	79	B	2.44	0.75	B	1.67	0.46	B	1.74	0.48
77		0.57	0.42	37		-0.31	0.46		0.20	0.30		0.15	0.31
78		-0.45	0.38	5		0.02	0.41		-0.22	0.27		-0.22	0.28
80		-0.97	0.89	83		-1.05	0.96		-1.02	0.65		-1.01	0.65
81		0.64	0.86	36		0.87	0.84		0.86	0.59		0.76	0.60
83		0.02	0.47	3	B	1.08	0.47		0.46	0.32		0.55	0.33
84		0.56	0.39	30		0.22	0.42		0.15	0.28		0.40	0.29
85		0.27	1.38	32		0.03	1.24		-0.10	0.93		0.14	0.92
86		0.13	0.64	92		-0.51	0.73		-0.11	0.47		-0.17	0.48
87		0.89	0.41	27		0.80	0.41		0.80	0.28		0.85	0.29
88		-1.04	0.71	87		-0.09	0.66		-0.74	0.48		-0.55	0.48
89		0.62	0.51	44		0.01	0.59		0.39	0.38		0.34	0.39
91		-0.64	0.47	24		-0.15	0.51		-0.24	0.34		-0.41	0.35
92		-0.06	0.62	1		0.32	0.65		0.19	0.44		0.13	0.45
93		-0.56	0.69	38	B	-1.67	0.69	B	-1.22	0.47	B	-1.12	0.49
94	B	-1.06	0.40	84		-0.76	0.42		-0.94	0.28		-0.91	0.29
		-----	-----			-----	-----		-----	-----		-----	-----
MEAN		-0.25	0.54			-0.25	0.55		-0.23	0.38		-0.23	0.38